Trajectory Clustering and an Application to Airspace Monitoring

Maxime Gariel, Ashok Srivastava and Eric Feron

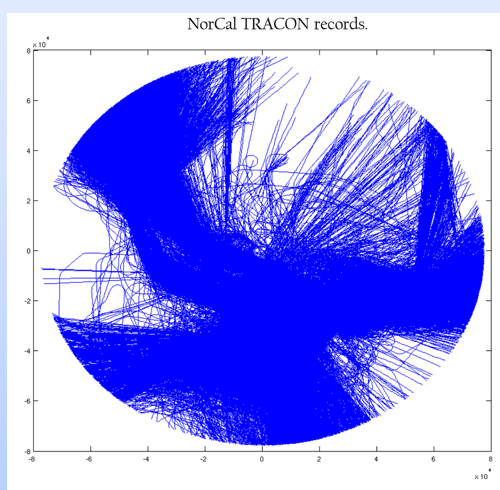
Abstract

This poster presents a data driven methodology to determine and learn “nominal” behavior of aircraft in the airspace surrounding major airports. To ensure the safety of the airspace and maximize runway occupancy, air traffic controllers (ATC) use standard procedures, guidance aircraft from a waypoint to another. Observed flight path present a large variability, and sometimes largely differ from standard procedure. Possible factors for this difference include: controllers allowing “direct routes”, skipping some waypoints in order to optimize the available resources, and weather. Another potential factor is that the control of the aircraft remains to the pilots.

In this work, we develop techniques to identify those procedures and their variability. We use recorded radar tracks from a terminal area. We present two trajectory clustering methodologies that enable us to obtain routes frequently flown. This knowledge base is then used to monitor the conformance of current to nominal operations. The objective is to monitor the instantaneous health of the airspace. We say that the airspace is healthy when all aircraft are flying according to the nominal procedures. When an aircraft does not, it requires more attention from the ATC and therefore diminishes the capacity of the controller.

Data Presentation

- Records of all flight tracks in the Northern California (NorCal) Terminal radar approach control (TRACON).
- Cylinder of radius 50 NM and height 20,000 ft centered at Oakland international airport.
- Data organized by flight. Contains metadata for each flight including: type of operation (departure/arrival), origin and destination airports, aircraft type, date and time of beginning of record, duration of the record.
- Focus on arrival flights to SFO 28R\L.
- Trajectories have different number of points (10 to 550)
- Interval between points depends on the rotational speed of the radar.



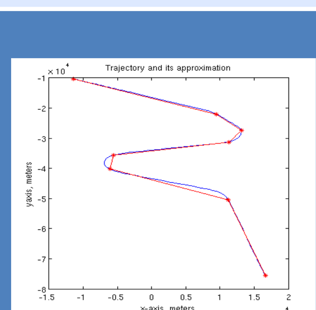
Trajectories to cluster
Over 27,600 arrivals on runways 28R/L at SFO

Waypoints Based Clustering

- Motivated by current published instrument flight rules procedures
- Determination of waypoints by finding and clustering the turning points in each trajectory.
- Clustering of trajectories that follow the same way points

Turning Points Identification

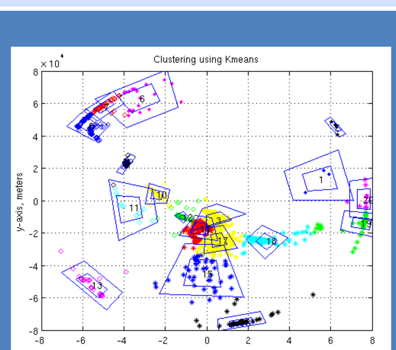
- Extraction the location of the turning points of each trajectory
  - Turning points locations = heading change
  - Heading estimated using data points
- Trajectories now represented as sequences of turning points



Sample trajectory and its approximation using turning points

Waypoints determination

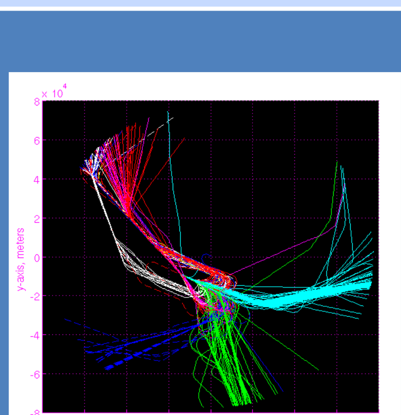
- Cluster the turning points using k-means.
- To each cluster corresponds a way point
- Trajectories now represented as a sequence of way points



Clusters of turning points for one day of data

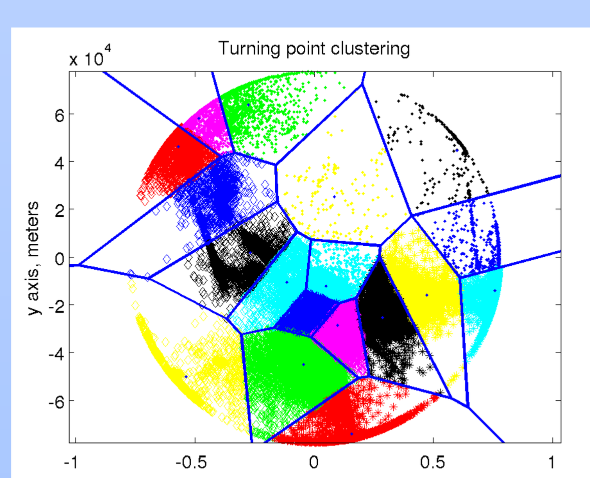
Determination of the longest common subsequence (LCSS)of waypoints

- Cluster the sequences of waypoints using LCSS.
- Allows to cluster sequences with different number of elements.
- SequenceMiner [1] identifies the (LCSS), generates clusters and provides centroids (representative trajectory for each cluster).



Clusters of trajectories for one day of data

Results and issues:
- Satisfying results for a limited number of trajectories.
- Identification of the waypoints is difficult and inaccurate with increased number of trajectories.
- Sequence of waypoints are short and the LCSS works better with long sequences.



Clusters of turning points for the entire data set

[1] S. Budalakoti, A. N. Srivastava, and M. E. Otey, *Anomaly Detection and Diagnosis Algorithms for Discrete Symbol Sequences with Applications to Airline Safety*, IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews, 2009

Maxime Gariel is a PhD student at the Georgia Institute of Technology
Ashok Srivasata is the Principal Investigator for the Integrated Vehicle Health Management research project at NASA
Eric Feron is the Dutton/Ducoffe professor of aerospace engineering at the Georgia Institute of Technology
This project was supported by NASA under Grant NNX08AY52A
This work was reali during the summer 2009

Full Trajectory Based Clustering

In this methodology, we do not assume that trajectories have an underlying structure. Trajectories are resampled with the same number of points, in order to use existing clustering algorithms such as k-means.

Resample the data

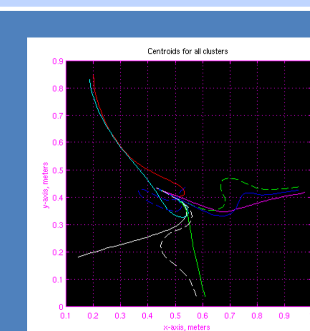
- Resample the trajectories, to obtain time series of equal length for each aircraft: 50 points.
- Discard trajectories with fewer than 50 points.

Preprocess the data

- Generate extra data such as: heading, airspace density, length of the trajectory, duration of the trajectory, etc.
- Add metadata to each flight, such as aircraft category, time of the day, etc.
- Normalize and concatenate all the data and metadata into a single vector for each flight

Cluster the data

- Cluster the data using k-means.
- Algorithm initialized with random trajectories and k large (30)
- Computation time < 1 min on a standard computer
- Convergence in 30 to 60 iterations
- Some clusters are empty

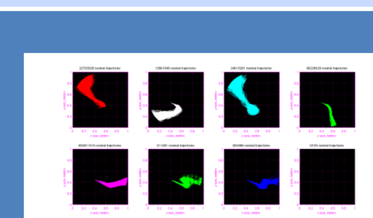


Cluster centroids

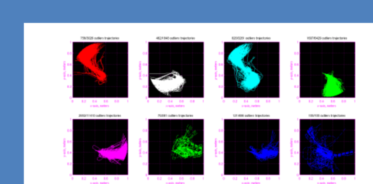
Clean the clusters

Identify and remove outliers

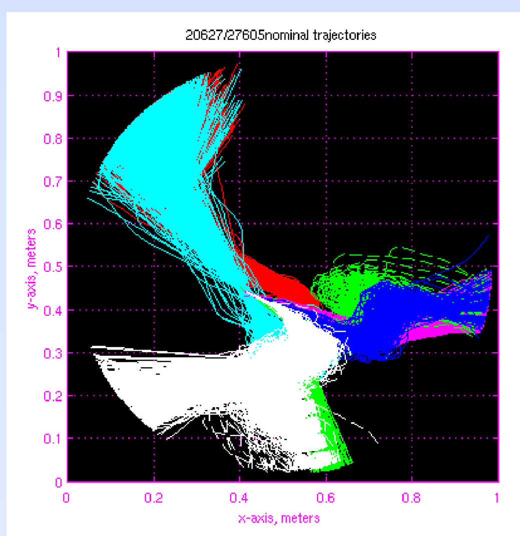
- Compute standard deviation of the cluster to determine clusters with too much variability: All the trajectories of a cluster with too much variability are tagged as outliers.
- Compare each trajectory with the centroid of the cluster: If they differ on too many points, the trajectory is tagged as outlier.



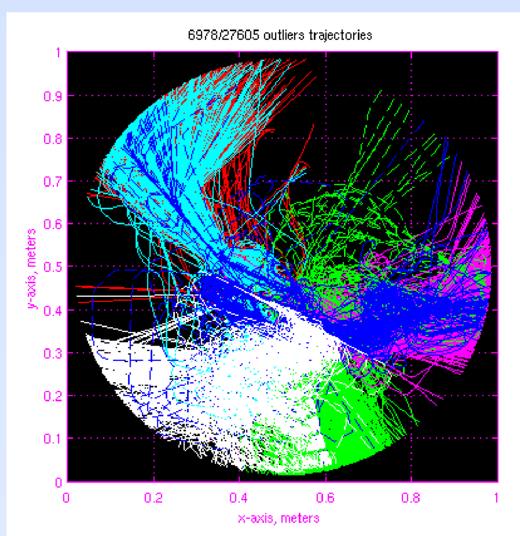
Nominal trajectories for each cluster



Outliers trajectories for each cluster



Trajectories identified as nominal



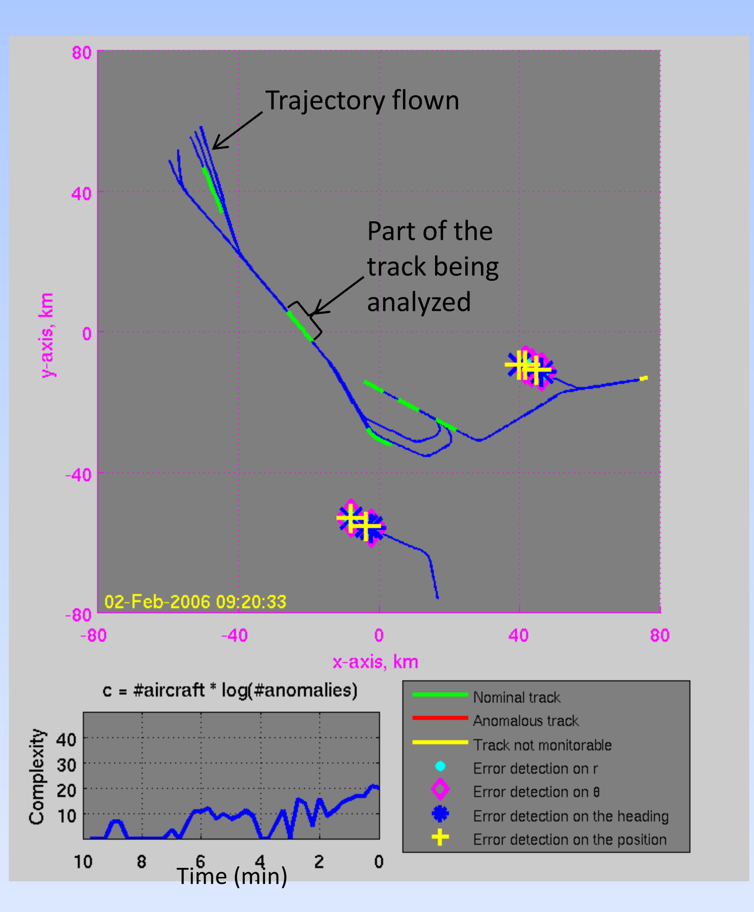
Trajectories identified as outliers

Results:
- This clustering and cleaning methodology provides results that are “visually” satisfying, since there exists no metric to assess the “correctness” of the results.
- Results depend on the initialization of the algorithm, but are consistent from one run to another.

Airspace Monitoring

Objective : identification of off-nominal trajectories in real-time

- Use the Inductive Monitoring System algorithm [2] (IMS) to determine the compliance of current aircraft trajectories to operations previously identified as nominal.
- IMS training data set is made of the nominal trajectories fragmented in vectors containing 5 data points.
- The last 80 seconds of record of current flights is resampled in order to be used as test data set for IMS.
- A trajectory is tagged as anomalous if the score provided by IMS exceeds a given threshold.
- The notion of time/speed is handled by the fragmentation.
- Measure of instantaneous complexity as:
  #aircraft *log(# anomalies)



[2] D. L. Iverson, and M. Stop, *Inductive system health monitoring*, Proceedings of The 2004 International Conference on Artificial Intelligence (IC-AI04)

Conclusion

This work presents two novel techniques for aircraft trajectory clustering. On one hand, waypoint clustering that shows good results on a limited number of trajectories. On the other hand, resampled full trajectory clustering that is able to handle a large number of trajectories and provides an easy way to clean the clusters. Nominal trajectories are identified after cleaning the clusters and used as a base line for identifying abnormal trajectories in pseudo real time. This works enables the monitoring of standard but variable operations: a flag is raised when an aircraft gets of tracks.